CDM Memorandum

Subject: Rationing Schemes for En-route Air Traffic Management

To: CDM CR Long Term Group (M. Ball, R. Beatty, A. Futer, R. Hoffman, K. Howard, B. Leber, M. Nadon, R. Oiesen, L. Sandusky, J. Sherry)

From: RBS-in-the-Sky Group (M. Ball, A. Futer, R. Hoffman, J. Sherry)

Date: February 22, 2002

Rationing Schemes

This is a summary of the various rationing schemes that were collected from members of the Long-term Collaborative Routing Working Group (LTCR). Each represents its own philosophy; none is worked out in complete detail. The most viable scheme(s) can only be determined through extensive testing, analysis and war game exercises. The ultimate rationing scheme for en-route resources may turn out to be a blend of the concepts expressed here.

The typical scenario is a demand-capacity imbalance for en route resources with sufficient notice to allow for implementation of an organized scheme. This could be certain types of weather disruptions, or simply too much demand. Some of the schemes spill over into the realm of tactical assignment of resources but, in general, they tend to be strategic in nature.

Whenever demand exceeds capacity for use of airspace, a decision must be made as to which flights will be rerouted (or delayed) by traffic flow management. It would be best if the carriers would draw down the demand themselves. So, it will be necessary to implement a constraint and demand notification system for en route resources. TFM and NAS Users would share such a tool. An official notification, or advisory, would be issued warning Users of those regions of airspace that require demand reduction. Users would be given a period of time to voluntarily draw down the demand over the time horizon of concern. If not enough flights are rerouted, then TFM must implement a formal procedure for rerouting flights.

There will be times when a user's best local strategy is to wait for a competitor to yield a constrained resource. The fairness algorithm must create a situation that makes the user's best local strategy to participate and adjust flight routes to alleviate constraints. Given a capacity vs. demand topology there may occur ridges of excess demand that will require delays either on the ground or due to circuitous reroutes. Allocation of who gets the good routes and delays must be fair. If non-participants receive the same or equivalent benefit the best user strategy is not to participate. Non-participants to preserve their competitive position will oppose implementation of any fairness algorithm that penalizes them for not spending dollars to participate.

Given the relatively short notice with which disruptive weather can materialize (e.g., one hour), some of these schemes may be more appropriate for over-demand situations. In all cases, however, we envision continuous NAS monitoring and flight planning, as opposed to on-off traffic flow initiatives.

We tried (but not always succeeded) to present rationing schemes below with general approach, operational concept, rationing mechanism, and benefits and disadvantages. The issues on which the schemes should be (but not always are) judged include:

- Efficiency: How fully the resources will be utilized, including NAS capacity, passenger delay (including down line one), etc.?
- Equity: How does the FAA decide how to spread the pain among existing traffic?
- Flexibility: How does a User make operational adjustments once the FAA has rationed the resources (that is, how does a User do equivalent of cancels and subs in a GDP)?
- Simplicity: How easy is it for FAA and NAS Users to comprehend the concept and naturally integrate it with current operations?
- Recoverability: If predictions (e.g. weather or demand) turn out to be wrong, how well does the system respond?
- Incentives: Is the User given incentives to participate in the scheme? [That is, we want to avoid situations where a User benefits who ignore or violate the rules. Put another way, what is the benefit for those who are good citizens and play by the rules?] Similar question: How easy it is to cheat within a scheme?

Several original documents were used to compose this one; they were subjected to some editing (including additional visualization) and placed into appendices. Hopefully, we did not through away babies with the bath water. Loraine Sandusky (Continental), Ken Howard (Volpe), Roger Beatty (AAL), late Mike Nadon (TWA), Bill Leber (NWA), Bob Hoffman (METRON), Mike Ball (UMD), Aron Futer (Volpe), Rick Oiesen (Volpe), Joe Sherry (MITRE) contributed to the document.

1. First-Filed, First-Served

In this scenario, a flight plan is seen as a contract for NAS resources between the FAA and the NAS User who submits it. An (accepted) flight plan is a 'reservation' for use of the NAS resources reflected in the flight plan. The FAA is bound to honor those reservations. In turn, the User must meet the specifications of the flight plan. The idea employs the concept of time/ location/ altitude to define both the User's intent and requirements in the deployment of NAS equipment and services, while at the same time, promoting filing of early but accurate intent. Alteration of the flight plan would be contingent upon availability and FAA approval.

The concept of a *controlled time of arrival* (CTA) is already in use. This is the time at which a flight is supposed to arrive at its destination airport, when under the influence of traffic flow management (TFM) initiatives. The CTA concept can be extended to each fix along a flight plan. By back calculating from a CTA, a *required time of arrival* (RTA) can be computed for each fix in a flight plan. This is the time at which the flight is required to pass over the fix in order to meet its CTA. The combination of CTA and a flight plan would determine a sequence of time-space ordered pairs: Fix_1, RTA_1, Fix_2, RTA_2, … Fix_n, RTA_n.

This embellished flight plan would be called a *Required Arrival Time / Controlled Arrival Time* (RATCAT). Of course, the major challenge is how to determine the RATCAT for each flight. Essentially, RATCAT would be based on a first-filed, first-served paradigm, in which the 'early bird' gets the best flight plan trajectory. This is not as simple as filing well in advance of (scheduled) departure because the best wind and weather data is generally not available until close to departure.

Computation of RTAs would require an advanced traffic projection (TP) module. Current TP modules input a sequence of 3-D points (lat/long/altitude) in the airspace that represent the path a flight plans to take. The output is a sequence of 4-D points, each of which is a 3-D spatial point along with a time at which the flight is predicted to be at the point. The advanced version would have to form the times based on desired arrival rather than departure. More importantly, the TP module would require advanced knowledge of NAS-wide traffic conditions to correctly compute the arrival times.

RATCAT would also require an interactive capability between Airline Operations Centers (AOC) flight planning computer systems and ETMS that allows Users to request a "reservation" and ETMS to reply if it is available.

See Appendix A for more details of this approach.

Benefits:
- improved flight planning due to taking into account knowledge about NAS constraints during trajectory modeling;
- improved predictability due to requirement to fly the filed flight plan;
- rewarding aircraft operators who correctly predict the route they will fly, which in part means rewarding those who have invested in advanced navigational technology.

Disadvantages:
- philosophy is first-filed first-served, taken too far. For example, "early and accurate intent" may become obsolete when the weather changes. Then the flights with early filing become disadvantaged, they are behind others with "reservations";
- low on recoverability;
- low on efficiency;
- does not allow for flexibility, e.g., what if flights spill into FCA from another troubled area;
- it is not clear how to apply other GDP-like mechanisms, as cancellations or substitution, their impact would be massive along the flight routes, and thus "reservations" would be violated;
- incites false filings (forced gaming for scarce resources). Actually, most of the schemes will be open for possible cheating;
- ETMS does not know what the sector capacities are (or even what FCAs are) several hours in advance;
- not clear how to deal with situations where the flight plan "contract" cannot be honored.

2. Accrued Delay as Uniform Equalizer

The underlying philosophy of this resource approach is that User intentions are best reflected by an original arrival time to destination. A flight may suffer from delays internal to its operation (e.g., mechanical delay) or from delays external to its operation, such as departure delays at the airport, miles-in-trail, delays caused by rerouting or other TFM initiatives. No matter what the delay source, it is in the best interest of all (especially the flying public) to help the flight minimize its arrival delay. For this reason, we establish the concept of net arrival delay (NAD), which is simply the difference of estimated arrival time at destination and original arrival time at destination:

NAD = (estimated arrival time at destination) – (original arrival time at destination)

For each flight, a pseudo-original time of arrival (POTA) can be constructed. POTA would be the difference of estimated arrival time at a sector or Flow Constrained Area (FCA) and net arrival delay:

POTA = (estimated arrival time at sector or FCA) – NAD

This way, when assigning virtual arrival slots to an FCA, flights can be ordered by increasing POTA, as in a GDP. In fact, the ration-by-schedule (RBS) algorithm used to allocate airport arrival slots in a GDP also minimizes the maximum delay. The difference is that the only delay considered in a GDP is assigned ground delay of the GDP itself. Under the proposed allocation scheme, delay is expanded to include all other forms of delay.

It is not yet clear how the "original" arrival time at destination will be determined. GDPs suffer from the same defect; many flights do not provide scheduled information. Perhaps the first message to enter the system would set the original arrival time.

It is relatively straightforward to compute the delays for a flight that is involved in multiple constraints. Independent impact assessment reports can be sent to the airlines for the multiple constraints. It would seem that because a flight can be involved in more than one constraint, the burden on the airline increases to determine the best ways to manage their schedules. However, with this burden may come additional benefits. For example, an airline might determine that by canceling a flight strategically placed in several different constraints, it could gain delay reductions for all flights involved in any one of these constraints.

Double penalty does not look to be an issue here. Suppose that a carrier reports a delay of Δ minutes to the FAA. The net arrival delay increases to $NAD^* = NAD + \Delta$, but ETA increases by the same amount, $ETA^* = ETA + \Delta$, thus POTA does not change.

See Appendix B for more details on this approach.

Benefits:

- tendency to reduce the total delay (efficiency of NAS usage, minimization of total passenger delay);
- simplicity; it is simple to understand this approach and would be relatively straightforward to implement;
- universality: it includes delay of any type: mechanical, miles-in-trail, GDPs, departure delays, delays due to rerouting, and so on;
- avoiding the double penalty issue;
- similarity to the GDP ground delay may allow for implementation of other GDP-like mechanisms (cancellations and substitutions);
- applicability to multiple FCAs.

Disadvantages:

- the same amount of accrued delay may have different value for long flights versus short flights; more generally, different flights may be of different importance for the Users; thus the claimed "uniform equalization" may not work out, especially if implementation of GDP-like mechanisms meets problems;
- the whole approach has a "Big Brother" flavor, with its pluses and minuses;
- the issue of different dwell times in the FCA is not addressed;
- doesn't distinguish rationing between flights running on schedule;
- leaves a lot of room for cheating (any approach likely will).

This approach may be expanded to incorporate downstream delay effects within an air carrier's operation. Late arrival of one flight may propagate delay to subsequent flights dependent upon the crew, equipment, or passengers. This can be measured by various means, such as down line passenger delay.

Originally, all forms of delay could be partitioned into three basic categories: air delay, ground delay, and downstream delay. Air delay and ground delay constitute NAD – net accrued delay. Adding down line delay will lead to defining a single delay metric, total imposed delay (TID):

TID = NAD + (down line delay) = (air delay) + (ground delay) + (down line delay)

This idea is outlined in more detail at the second section of Appendix B, which is named 'Intent-Neutral Total Imposed Delay (INTID)'. Actually, additional considerations associated with down line passenger delay would enrich other rationing schemes as well. Those of the schemes that would not allow it are disadvantaged.

Note. The whole approach has a "Big Brother" flavor, with its pluses and minuses.

3. Reroute Roulette

In a flow-constrained situation, a certain percent of the traffic must be displaced to accommodate decreased capacity. Ideally, NAS Users should be made aware of the constraints and given the opportunity to reroute around the flow constrained area. If not enough flights were moved, then a random selection, or lottery, process would perform traffic displacement. A list of flights intending to fly through the FCA would be formed and flights would be randomly selected from the list and rerouted until demand is in line with capacity. Those Users who voluntarily rerouted around the flow-constrained area would receive some special consideration in the lottery.

Equitable treatment of NAS Users is a prime concern for an TFM initiative. Under this proposed rationing scheme, a random selection process would maintain equitable treatment of flights. For example, suppose that a demand exceeds capacity for a fixed region of airspace. Then for each time period t of concern, X_t flights must be moved to reduce demand to the capacity level during period t . A list of flights intending to use the region would be formed, and flights will be selected one at a time from the list until X_t flights have been chosen. These flights would then be rerouted around the flow-constrained area.

Slight inequities may arise in a given implementation of this "reroute roulette", but these will be smoothed out with more frequent use of this rationing concept. In the long run, flight displacement will be in proportion to a carrier propensity to use flow-constrained areas.

In order to provide an incentive for Users to alleviate congested situations without resort to TFM initiatives, special consideration must be given to those Users who reroute flights in response to FCA advisories. This can be accomplished by granting a 'free draw' in this lottery process for each flight moved. Suppose that carrier XAL has voluntarily rerouted a flight around an FCA. Then the first time that an XAL flight is drawn from the selection pool, the flight will be returned to the pool. Similarly, if XAL voluntarily rerouted 10 flights, then the first ten times an XAL flights is drawn from the pool, the flight would be returned.

Under the current filing system, Users would retain the ability to subvert the random reroute process by re-filing through the flow-constrained area. Therefore, it would be necessary to ban the selected flights from use of the resource for a sufficient period of time (e.g., 24 hours after the advisory is issued).

Slightly more detailed description of this approach may be found in Appendix C.

Benefits:

- this encourages early filing of intentions;
- users have a benefit if they voluntarily reroute around a constraint;
- the decision as to rerouting around a constraint is left to the user's risk to benefit analysis of the individual situation;
- most resource allocation problems are resolved before departure. Users can detect delays and work on down line solutions sooner which gives more solution options;
- en-route flights are included in resource problem resolution;
- changes in the NAS while flights are en-route can be handled in the same manner with adequate lead time;
- the schema is relatively simple and easy to implement;
- equal treatment of NAS Users (more or less).

Disadvantages:

- until the time limit on the process is reached a user is subject to uncertainty as to the route to be flown since each new flight overloading a resource results in another lottery;
- in the extreme case a flight can lose enough lotteries that it becomes impossible to operate the flight that day;
- the penalty for a flight which is operating at the edge of its specific range can be diversion. The user has no alternate route they can fly without an en-route stop so voluntary rerouting is not an option. If the flight loses the lottery an unplanned landing will result;
- does not address temporal delay, only spatial delay is addressed;
- implementation of GDP-type mechanisms is not defined;
- application of this approach to multiple FCAs is not defined yet, may be complicated;
- part 129 carriers will be required to participate which has treaty implications.

Issues:

- Will this work for carriers with very small numbers of flights?
- Are we sacrificing optimality by selecting flights at random? May be the flights are not that equal after all? [Counter issue: for those who wary about optimality – see the next section.]

4. Big Brother

The philosophy of this approach would be global optimization by a central decision-maker. The system will resolve congestion by rerouting and/or delaying some aircraft. The aircraft to be rerouted or delayed will be picked up from a pool of aircraft, based on some global criteria. These criteria will most likely include a wide variety of factors, depending on the scope of a congestion problem. This could be an isolated and rather limited in size flow-constrained area, like a local thunderstorm area or an overloaded sector; or, at the opposite, massive traffic disruption caused by a huge severe weather activity; or any option in between.

There are different ways to consider multiple factors that would impact the decision-making process. For instance, it may be a hierarchical system where various factors are taken into consideration one at a time in some order; or these factors could be combined with some weights into a single goal function. In either case, there will be some predefined and mutually agreed mechanism between FAA and the NAS Users on how to build the hierarchy and assign the weights. And then everybody agrees to abide by the decision suggested by the optimization software.

Unfortunately, rather than some input in the optimization function development and fine tuning not a lot of involvement is seen for NAS Users. So it is unlikely to be implemented as a single tool; it could be one of the tools though. On the other hand, equal treatment of Users will be achieved. The decisions to be made do not depend on air carrier and thus will be able to smooth out possible unevenness (on average).

A product had been developed at Volpe Center that could have been used in accord with this philosophy – the ETMS Rerouter; it was not deployed though. The Rerouter was capable of finding optimal routes for aircraft around areas of airspace impacted by severe weather with consideration of sector loads. It used a one-dimensional goal function with tunable weights assigned to various factors.

A version of implementing the Big Brother approach may go along the following scenario.

Air carriers send schedule information 24 hours before departure. 18 hours before departure ATC uses the current wind model to create routes for each IFR flight that avoids all known and forecast constraints and balances resource demand on the NAS. These routes are forwarded to the users and the user is awarded slots at each resource based on ATC selected routings. Resources that are available but not allocated are also available for use on a first come, first served basis after the initial route awards are complete.

Users plan their operations using these routes and when appropriate exchange resource slots among their flights to meet their operational needs. ATC and users would collaborate on the users special case needs.

Non-participants would have to plan their flights to avoid all constrained resources.

As changes occur in the NAS and resource availability changes users will be advised of slots that open or close.

Benefits:

- you can explicitly state and achieve the goals, like overall reduction of total delay or fuel consumption;
- global optimization may provide system-wide benefits;
- hierarchy of factors and/or assigned weights will allow to flexibly manipulate and fine tune the software system (if implemented properly);
- equal treatment of NAS Users (on average) - a neutral does the initial allocation of resources;
- avoiding double penalty (or minimizing maximum delay).
- non-participants are sufficiently penalized that few users will choose not to participate;
- Users can avoid most ground delays by appropriate allocation of resources among their flights;
- most resource allocation problems are resolved before departure. Users can detect delays and work on down line solutions sooner which gives more solution options;

Disadvantages:

- the approach is "Big Brother", taken too far. In the worst case due to variations in flight planning data and parameters the resources allocated to a user may be totally inappropriate to their operational needs. (E.g., a User operating hourly flights between MCI and BOS may plan flights at maximum speed and below FL 280 to minimize block times. Winds and allocated resources used by ATC may be for FL 310 and above for this city pair leading to different routings and resources than desired by the user);

- implementation of GDP-type mechanisms may require sacrificing some global optimality to allow cancellations and substitutions; [may be this is benefit, not a disadvantage?];
- GA and on demand users will be more constrained than scheduled carriers;
- as changes occur in the NAS (TRW, outages, GDP etc.) the resources allocated to a user may be constrained more than forecast requiring them to accept sub optimal routes. Changes to the NAS may also result in loosened constraints at a resource allowing more optimal routings.

5. Traffic Classes

A general philosophical approach to traffic flow management and rationing involves the use of aggregate traffic classes. The use of traffic classes provides a hierarchical approach to problem solving that is consistent with many decision support tools and perhaps, more importantly, with the way traffic flow managers approach problem solving. For example, traffic flow managers might speak in terms of northeast arrivals, departures to the west, over-flights, etc. Since any automatic algorithms for controlling flights and allocating resources must be monitored and controlled by human decision makers using an approach that is consistent with their intuition has many advantages.

We define a *traffic class* as a set of flights with certain common characteristics. The common characteristic(s) could be a route of travel, a destination point (airport, departure fix), or geographical description (e.g., eastbound flights). A flight may be in more than one traffic class. For instance, a flight could be in the eastbound traffic class and in the EWR arrival traffic class. Each flight must be assigned to at least one traffic class. Establishing an "other" (default) traffic class can insure this.

We envision that some of the traffic classes would be established automatically by route bundling or by historical traffic patterns. The user would reserve the right to delete a traffic class or to create a new one. One of the most fundamental traffic classes would be those flights that are scheduled or predicted to pass through a flow constrained area (FCA).

Traffic classes serve four purposes. They allow the user to

- aggregate a large number of flights into a manageable number of groups for the benefit of discussion and conceptualization of flows;
- insure equitable treatment for flights through explicit representation (e.g., departures attempting to merge into an overhead stream);
- control flows by entering restrictions or capacities on each traffic class;
- reroute flights on a large scale by moving flights from one traffic class to another.

This concept is deliberately general/flexible, and is designed to aggregate flights much the same way that traffic managers do when thinking about traffic flows.

Of particular significance here is that traffic classes provide an approach to rationing resources and to considering equity in a manner that would otherwise be difficult. Associating aggregate traffic flow rates with each traffic class would perform resource allocation. These flow rates would serve as capacities in a second allocation step which assigned routes and times to individual flights. For, example one might assign a rate of 20 flights per hour with the Detroit departure traffic class (for a certain time period), another rate for east to west over-flights, another rate for west to each over-flights, etc. By doing this, traffic flow managers could insure a fair balance among departures, over-flights and other traffic. Analytic tools would, of course, be required to support the traffic flow manager in setting these rates.

Once flow rates/capacities were assigned to traffic classes, a second level of rationing procedures would be required to set routes and timing for individual flights. If a flight were in a constrained traffic class, then there would be two options for relieving the excess demand:

1. reroute (reassign to another traffic class);
2. ground delay.

This will create conflicts for resources that must be resolved. In particular, this must be done in a way that satisfies the flow constraints on all of the traffic classes. This level of resource allocation should make use of many of the CDM concepts developed for GDPs. In particular, resources would be initially allocated to individual flights, but then airlines would have the opportunity to exchange the assignment among its own flights. This is an area that clearly requires R&D attention. A flow balancing optimization problem (integer program), which to carry out the allocation step, is described in Appendix D.

Benefits:
- follows current pattern of thinking of traffic managers;
- provides for equal treatment of NAS Users.

Disadvantages:
- new fairness concepts must be defined and accepted;
- implementation of GDP-type mechanisms are not defined yet.

6. Near-Term Weather Avoidance[1]

This scheme is "Near-Term" in two senses: The scheme is intended to be capable of being fielded in the near future. Also, the scheme is primarily designed for the tactical TFM environment where weather forecasts are accurate enough to assure that flights will want to avoid the weather, typically within a two-hour forecast window.

The Near-Term scheme has potential benefits as an interim approach for better TFM decisions while researchers develop advanced collaborative systems. Advanced automated collaboration capabilities that would allow dispatchers to work towards custom preferred solutions for each flight can eventually be incorporated into this scheme, or could even replace this scheme.

The major goals that this scheme attempts to achieve are to:
- Keep sector workloads under Monitor Alert Parameter levels;
- Ground delay inactive flights until the automation reroutes active flights;
- Minimize the arrival delay for each rerouted or delayed flight;
- Provide quick solutions to allow an operator to perform multiple what-if evaluations using different parameters and routes; and
- Avoid bias towards any NAS user or traffic class.

In this scheme, weather FCAs are automatically (or manually) created using automated weather forecasts. Next, operators manually select reroute corridors to avoid this forecasted weather. The user community may predefine these reroute corridors or operators may dynamically create them. Then, operators determine the list of flights for the automation to reroute. Automation helps in this process by providing a list of those aircraft whose flight paths will intersect the predicted weather. The selected flight list may simply be the list provided by the automation or operators may add to or remove from this list. Next, an automated assignment algorithm assigns each flight to a reroute corridor using the goals described earlier and the following steps:

1. Reroute active flights that are associated with the weather problem;
   - Use 1st come 1st serve (i.e., earliest arrival at best reroute corridor option for each flight) to determine which active flights to reroute first;
   - Of the reroute options available, select the option with the least reroute delay that does not exceed the Monitor Alert Parameter (MAP) thresholds for the sectors involved in the reroute; and
   - If all options exceed MAP thresholds, select the reroute involving the least sector overload.
2. If MAP thresholds are exceeded in Step 1, ground delay as necessary inactive flights that are contributing to the congestion problem (these are <u>not</u> flights associated with a weather problem).
3. If MAP thresholds are <u>still</u> exceeded from Step 2, apply airborne delay to aircraft with a weather problem to resolve remaining sector overload.
4. Reroute inactive flights associated with the weather problem and ground delay them if necessary;
   - Use 1st come 1st serve to determine which in active flights to reroute first;
   - Select the reroute option with the least combined reroute and ground delay that does not exceed MAP thresholds for the sectors involved in the reroute; and
   - If all available options exceed MAP thresholds, select the reroute involving the least combined reroute and ground delay.

Next, automation provides an assessment of the plan. Lastly, users review and modify the plan as desired.

Benefits:
- can be used as an interim approach while advanced collaborative concepts are developed;
- provides a baseline to measure other equitable allocation schemes or combinations of schemes against;
- allows the complex issue of NAS efficiency and user equitability to be studied and assessed in a more realistic setting;
- utilizes sector loading limit thresholds, so sector congestion problems can be identified and simultaneously resolved.

Disadvantages:
- has no automated features for collaboration and handling dispatcher or airline preferences, however work is in progress to include collaborative capabilities;
- is only equitable to the extent that the system does not consider flight Ids during evaluation. But, other equitability schemes could be added as assessments identify user or flow inequities.

Appendix F describes the environment of the related work in more detail.

7. A New Ground Delay Program

Use ground delays to alleviate constraint problems. When the user files a proposed 4D trajectory the route is checked for any sector constraints. If there is a constraint the ATC system advises the user of the delay along this 4D trajectory. The flight will be delayed until it will not overload any facility. Maximum time from proposed departure a user may file should be established such as no filing more than 12 hours before departure. When a delay is imposed the user receives a delay message which includes the explanation of the constraint. The user may then test another possible route against the constraint system. "New Age" flight plan filers could add a 2nd and 3rd route with associated acceptable delay. The ATC filing system would accept alternatives based on delay parameters if the primary route was not available at the proposed time.

Benefits:

- this encourages early filing of intentions;
- non-participants are nonexistent since any IFR flight plan filer is in the process;
- plans can be amended and tested against the ATC constraint system to obtain the best trajectory consistent with their needs. Users could automate much of this process;
- operators with multiple flights through a constrained area can choose which flights to consistent with their needs. Once a route is awarded each user has a series of CTA's at various points en-route that they can swap among their flights;
- most resource allocation problems are resolved before departure. Users can detect delays and work on down line solutions sooner which gives more solution options;
- users could automate alternate route selection criteria in accordance with their safety and business needs.

Disadvantages:

- requires creation of a huge database with difficult functional requirements. Automating the checking of proposed routes and preserving user CTA's along a route is required;
- once flights are airborne this approach doesn't address changes in NAS Status;
- flights delayed for maintenance or other reasons could go through a difficult and repetitive process with each new ETD yielding new constraints and route changes that could add to the flight's delay;
- as in GDP cheating is possible but more difficult to detect. A user trying to preserve slots in congested airspace could file all their transcon flights through the area twelve hours before they depart. When the twelve-hour threshold is reached the user could file for this area with shorter haul flights and use the transcon CTA's to swap with. Thus getting the best route for their short hauls and putting their long hauls on their actual preferred routes;
- Part 129 carriers will be required to participate which has treaty implications.

8. Collaborate or Die

Twenty-four (24) hours before the proposed departure the flight's UPT (user preferred trajectory) is filed with ATC. The flight plan filing includes an acceptable alternate routing and a maximum delay acceptable before the user will accept the alternate route. ATC processes these requests twelve hours before schedule departure and allocates routes to the users based on their filing.

ATC sends advisories to users about constraints in the system that are exceeded and what proportion of flights each user must move from which constraint. This proportion is based on the proportion of that user's flights the user has scheduled to use that resource. If the user fails to remove enough demand then the earliest departing flight(s) for that user to that resource is delayed until the demand equation for the resource is acceptable. The user is then able to swap their remaining slots for that resource among their flights and/or reroute the flights that are delayed.

Users with a single flight through the constrained resource are all placed in a common pool and the earliest departing flight(s) are delayed until the resource can accept them. The user can accept the delay or amend the route.

As a resource is overloaded and delays are imposed the time frame for the resource overload may grow. The solution set for any "En-route delay program" should include all flights in the expected time frame. This will allow more options to more users.

Non-participants who file as they currently do will be given delays if their route intersects any constraint. ATC will not be expected to provide non-participants with alternative routes.

Benefits:

- non-participants are sufficiently penalized that few users will choose not to participate;
- Users can avoid most ground delays by appropriate allocation of routes among their flights;
- most resource allocation problems are resolved before departure. Users can detect delays and work on down line solutions sooner which gives more solution options;
- alternate flight route filings give the user more opportunity to meet their operational and business goals.

Disadvantage:

- RBS may favor larger operators since single flights in a program can receive extensive delays;
- short haul operators may be penalized more than long haul operators;
- Part 129 carriers will be required to participate which has treaty implications.

Appendix A.

Required Arrival Time Controlled Arrival Time (RATCAT)

Philosophy:

A flight plan is a contract for NAS resources between the FAA and the NAS User who submits it. An (accepted) flight plan is a 'reservation' for use of the NAS resources reflected in the flight plan. The FAA is bound to honor those reservations. In turn, the User must meet the specifications of the flight plan. RATCAT employs the concept of time/ location/altitude to define both the User's intent and requirements in the deployment of NAS equipment and services, while at the same time, promoting filing of early but accurate intent. Alteration of the flight plan would be contingent upon availability and FAA approval.

Concept:

The concept of a *controlled time of arrival* (CTA) is already in use. This is the time at which a flight is supposed to arrive at its destination airport, when under the influence of traffic flow management (TFM) initiatives. The CTA concept can be extended to each fix along a flight plan. By back-calculating from a CTA, a *required time of arrival* (RTA) can be computed for each fix in a flight plan. This is the time at which the flight is required to pass over the fix in order to meet its CTA. The combination of CTA and a flight plan would determine a sequence of time-space ordered pairs, as in Figure #.

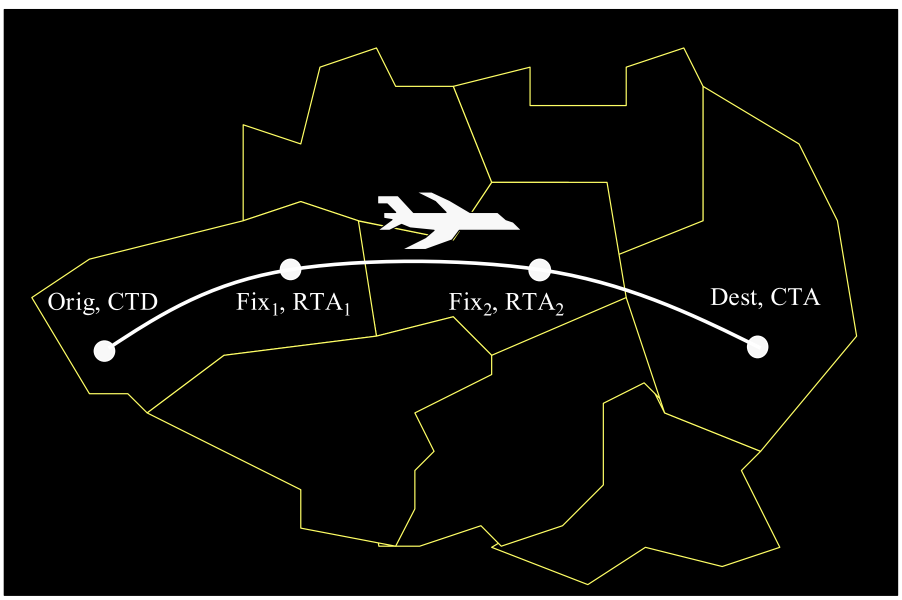

Figure A: RatCat Concept.

This embellished flight plan would be called a *Required Arrival Time Controlled Arrival Time* (RATCAT). RATCATs can be used in a self-imposed rationing scheme. The RTA waypoints would be the existing VOR, FRD, or lat/long points. RAT parameters would be calculated based on projected demand in the airspace, as well as aircraft performance and payload. During peak periods, the RTA in congested airspace might be configured to plus or minus X minutes; while night flights on the same trajectory might be issued a RAT of plus or minus Y minutes. A subset might be different RTAs along the same route based on different levels of congestion modeled by traffic managers in that local airspace to prevent

Mile In Trail (MIT) restrictions. Hence, the aircraft FMS [acro?] would be speeding up or slowing down to meet the flight planned RTAs along the route.

The proposed philosophy is that a RATCAT becomes a contract for NAS resources between the FAA and the NAS User who submits it. If the RATCAT is accepted by the FAA, then the User has in essence ‘reserved’ use of the NAS resources reflected in the RATCAT. The FAA is bound to honor those reservations. In turn, the User must meet the specifications of the RATCAT. Alteration of the flight plan is contingent upon availability and FAA approval, and would probably result in a weighted penalty.

Of course, the major challenge is how to determine the RATCAT for each flight. Essentially, RATCAT would be based on a first-filed, first-served paradigm, in which the ‘early bird’ gets the best flight plan trajectory. This is not as simple as filing well in advance of (scheduled) departure because the best wind and weather data is generally not available until close to departure.

Computation of RTAs would require an advanced traffic projection (TP) module. Current TP modules input a sequence of 3-D points (lat/long/altitude) in the airspace that represent the path a flight plans to take. The output is a sequence of 4-D points, each of which is a 3-D spatial point along with a time at which the flight is predicted to be at the point. The advanced version would have to form the times based on desired arrival rather than departure. More importantly, the TP module would require advanced knowledge of NAS-wide traffic conditions to correctly compute the arrival times.

RATCAT would also require an interactive capability between Airline Operations Centers (AOC) flight planning computer systems and a next generation Traffic Situation Display (TSD), replete with predictive Flow Constrained Area (FCA) algorithms and conflict probe technology. Incorporating predictive and real-time conflict probe, wind, weather, and airspace constraint technologies would allow both airspace Users and Traffic Managers to equitably ration airspace resources based on “first filed–first served”.

RBS-type adjustments would have to be made for issues like short haul vs. long haul flights, as well as for non-scheduled (GA) operators.

Flights with a stage length exceeding two hours would be able to file optional trajectories (as in NMPS airspace today). These optional routes might be based on different wheels up times, or different aircraft weight (once the flight is closed out with weight and balance calculated). Using concepts in the National Playbook, these might be based on avoiding weather at departure, enroute, or arrival environments.

One of the primary benefits of RATCAT is improved flight planning. Currently, all NAS Users plan and file “in the blind”, without knowing NAS constraints during trajectory modeling. Once the dispatcher selects an optimum trajectory, it would be permanently filed with an actual P-time. All flights, regardless of stage length, would receive a Controlled Arrival Time (CAT) for the destination runway, along with a set of Required Times of Arrival (RTA) for each lat/long/alt point on the flight plan.

Another benefit is improved predictability. Currently, pilots and controllers dynamically alter the planned trajectory with (unintended) down line consequences of exacerbating congested chokepoints. In RATCAT rationing, pilot-controller directs would wreak havoc on another aircraft’s required time at a particular latitude/ longitude/ altitude fix. Hence, flying the filed plan would give airspace operators as well as airspace providers, the predictability so elusive in today's NAS.

Another benefit of RATCAT is that it rewards aircraft operators who have invested heavily in next generation aircraft to realize an ROI (return on investment) allowing them to utilize onboard CNS (communications navigation surveillance) technologies to minimize delays by flying flight plans that optimize their corporate business objectives.

Issues:

- A traffic congestion / flight planning query database (FPDB) would allow Users to query best trajectory options between point A and B, based on weather, weight, speed, (cost indices?), preferred arrival time at the destination airport, and projected congestion. State of the art flight planning systems would incorporate all available predicted and real-time NAS data streams to calculate best trajectory options. Lesser systems not possessing the ability to crunch all available data into the most viable options would still yield more quality flight planning and filing than we have today.
  The problem with this though might be that this is too complex a query. The query should be that an airline sends a route to ETMS and ETMS tells it if it can be flown without overloading any resource. If ETMS O.K.s the route, then the contract is signed. It is up to the airline to weigh all the factors (distance, etc.) and to evaluate different routes.
- Conflict probe algorithms would track how many flights have been filed at a particular lat/long/alt point. Basic separation would be achieved by disallowing another flight to file over the same point in space within a specified interval of time. This would minimize use of mile-in-trail (MIT) restrictions.
- Current pilot-controller directs would have to be curtailed to safety-only issuances, in order to produce the necessary predictability.
- Advanced navigation technologies in the cockpit would be utilized to support tactical precision flying in support of strategic predictability in the National Airspace. RNAV technologies would be used very precisely to enhance capacity and throughput in the NAS. Dynamic flight planning and trajectory adjustment would be made while an aircraft is en route to take advantage of unused (unoccupied) airspace arising from mis-forecasted demand or capacity.
- Airborne-wise CNS technologies aboard the aircraft would be used to support precise flying time to a fix as well as down linking RAT data via CPDLC type interaction.
- National Airspace grid model using latitude, longitude, and altitude waypoints overlaid with current (and future) airspace design. Cubic lat/long/alt points would be interwoven into static PDR/PAR traffic flows. Enroute, lat/long/alt points would replace railroad track airways. RNP and RVSM values would also be factored into the enroute cubic grid. Transition environments using discrete lat/long/alt points could be created to handle ingress/ egress traffic from major hubs. Transition airspace would flow into RNAV arrival/ departure procedures, which might encompass several airports to harmonize and optimize runway configurations.

Appendix B.

Equity via Net Arrival Delay (ENAD)

Philosophy:

Equity, rationing concepts, and lessons learned can be extrapolated from the GDP setting. For rationing en route resources, priority would be given to flights with the greatest net arrival delay (NAD), which is the estimated arrival time minus the original (scheduled) arrival time. NAD captures the net effects of any (forecasted) disruptions to the arrival of the flight, such as miles-in-trail, departure delays, mechanical delays, or other ground delays.

Concept:

Under this allocation scheme, virtual arrival slots to a flow constrained area (FCA) would be created according to capacity forecasts, much like in a GDP. A list of flights bound for the FCA would be generated and ordered by a priority scheme; the flight at the top of the list would be assigned to the earliest possible arrival slot and deleted from the list in a recursive manner.

The question is how to prioritize the flights. In a GDP, flights are ordered by original arrival time to the afflicted airport. This concept does not carry over to an FCA because not all flights originally intended to pass through the FCA. However, one can create a pseudo-original time of arrival (POTA) for each flight by back computing from the original arrival time at destination. Of course, the equity of this scheme will vary highly with the manner in which this is done.

The underlying philosophy of this resource allocation scheme is that User intentions are best reflected by an original arrival time to destination. A flight may suffer from delays internal to its operation (e.g., mechanical delay) or from delays external to its operation, such as departure delays at the airport, miles-in-trail, or other TFM initiatives. No matter what the delay source, it is in the best interest of all (especially the flying public) to help the flight achieve its intended arrival delay. For this reason, we establish the concept of net arrival delay (NAD), which is simply the difference of estimated arrival time at destination and original arrival time at destination:

NAD = (estimated arrival time at destination) – (original arrival time at destination)

Note that NAD includes delay of any type: mechanical, miles-in-trail, GDPs, departure delays, and so on. Any disruption to a flight achieving its intended arrival time is reflected in the NAD.

For each flight, a pseudo-original time of arrival (POTA) can be constructed. POTA would be the difference of estimated arrival time at the FCA and net arrival delay:

POTA = (estimated arrival time at FCA) – NAD

This way, when assigning virtual arrival slots to an FCA, flights can be ordered by increasing POTA, as in a GDP,. This has the effect of smoothing out and equalizing delays over all flights. Equity is achieved via net arrival delay (ENAD).

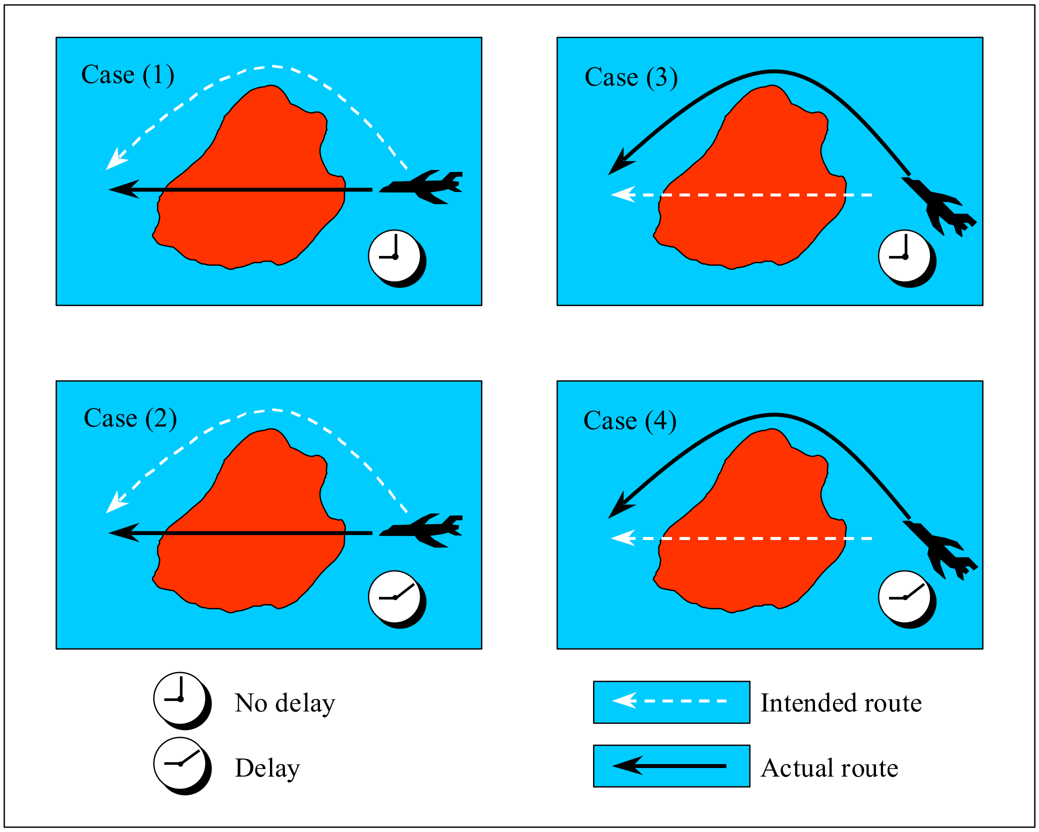

Figure B: Net Arrival Delay.

Let's consider the consequences of ENAD. For a given flight *f*, there are four cases with respect to computation of POTA:

(1) *f* was originally scheduled to pass through the FCA and incurs no additional delays
(2) *f* was originally scheduled to pass through the FCA and incurs additional delays
(3) *f* was not originally scheduled to pass through the FCA and incurs no additional delays
(4) *f* was not originally scheduled to pass through the FCA and incurs additional delays

The four cases are depicted in Figure #. In cases (1) and (3), POTA will be (roughly) the time at which *f* would normally pass through the FCA without a related TFM initiative. Case (3) could arise if a flight were rerouted into the FCA. In cases (2) and (4), POTA would be the time at which *f* would pass through the FCA, had it not incurred any other delays.

If a flight of type (1) were competing with a flight of type (2) for the same arrival slot to the FCA, then priority would be given to the type (2) flight (the delayed flight). Similarly, priority would be given to a type (4) flight over a type (3) flight. In essence, no consideration is given to early intention of passing through the FCA; the only factor taken into account is the amount of NAD.

You might think that a diverted flight would require special processing. However, a diverted flight still has an original arrival time at the destination, which would be taken into account when setting the POTA of the recovery leg. Say, for instance, the recovery flight is due to arrive four hours beyond the original arrival time of the diverted flight leg. Another flight bound for the same airport did not divert but has been sitting on the ground for six hours ready to depart. Under this scheme, the grounded flight would be allowed to go first, since it has accrued more total delay.

It is worth noting that an equivalent allocation scheme is to

> *make an assignment of flights to virtual arrival slots that minimizes the maximum net arrival delay over all flights.*

In fact, the ration-by-schedule (RBS) algorithm used to allocate airport arrival slots in a GDP also minimizes the maximum delay. The difference is that the only delay considered in a GDP is assigned ground delay of the GDP itself. Under the proposed allocation scheme, delay is expanded to include all other forms of delay.

Whenever the ETA of a flight changes, including from a traffic management initiative, a new ETA would be modeled and both the NAD and the POTA would be recomputed. To see that this avoids the double penalty issue, suppose that a carrier reports a delay of Δ minutes to the FAA. The net arrival delay increases to $NAD^* = NAD + \Delta$, but ETA increases by the same amount, $ETA^* = ETA + \Delta$. The new POTA becomes:

$$POTA^* = ETA^* - NAD^* = (ETA + \Delta) - (NAD + \Delta) = ETA - NAD$$

Which is the same as the POTA. There are no negative ramifications from having reported the delay.

Recall that NAD is the difference of ETA and original arrival time. It is not yet clear how the "original" arrival time at destination will be determined. GDPs suffer from same defect; many flights do not provide scheduled information. Perhaps the first flight message to enter the system would set the original arrival time.

At first glance, it may appear that this scheme does not allow for spatial delays, that is, rerouting of aircraft. However, we leave this option to the carriers. We envision that prior to issuing ground delays for an en route constraint, the FAA would notify the carriers of

- the constraint
- the affected flights (forecasted)
- delays that would be incurred, if no action were taken by the carriers.

The carriers would then have the opportunity to reroute, if the ground delays were unacceptable to them. Ground delays would be issued only if still needed. It may be possible to issue a credit to carriers that voluntarily reroute around the FCA.

A ramification of this approach is that we would not distinguish traffic management delays from other delays. So, if a flight has a one-hour mechanical delay, it would get the same priority as a flight with a one hour GDP delay. The arguments here are:

- Isn't it to everyone's benefit to have the system trying to keep all flights on schedule regardless of the reason for its delay? (This seems to be acceptable in the GDP.)
- Is it worth the added complexity to handle it any other way?

So in summary:

- When a flight is first created, we save the original arrival time.
- Whenever there is a change to a flight, including a planned traffic management action that will affect that flight, we model the new ETA and compute the accrued delay.
- Whenever a subset of flights must be selected from an otherwise acceptable for a traffic management action group (e.g., these 10 flights could all fly this other route, we need to pick three), we pick the flights with the least accrued delay.
- When we are assigning delay (e.g., slots) to flights, we do so in a manner such that the accrued delay of all the affected flights will be as equal as possible after the assignment is complete (like in a GDP).

Now, how does a user make operational adjustments once the FAA has rationed the resources (that is, how does a user do the equivalent of cancels and subs in a GDP)? This problem is much more complex. Again, let us start with the tried and true GDP and see to what degree we can extrapolate it to the whole system. These are the particular elements of the GDP process that make it successful from the airline perspective:

- An airline can see exactly what the expected impact is on its flights, and therefore on its schedule

- An airline can manage its schedule in two ways:
    - it has the ability to shift delays from one flight to another
    - it can reduce the delay on some of its flights at the expense of canceling other flights

From the FAA perspective, this flexibility offers the following advantage:

- An airline has motivation to reduce its load on the constrained area.

- The FAA can ultimately resolve the problem by just issuing the ground delays.

How can we preserve these key elements in en route airspace? Let's start by considering a single constraint – an overloaded sector. Here is how a sector problem could be resolved:

- When a constraint is first identified, the FAA provides the airline a list of the impacted flights and how much delay they can expect given the current demand and routing. (The delay would be equitably assigned based on the previous equity concept.)

- The airline assesses whether the delays are acceptable or not.

- If the delays are not acceptable, the airline considers its options:
    - The airline can transfer delay between flights as in GDP subbing.
    - The airline can cancel flights and move other flights up as in GDP subbing.
    - The airline can voluntarily route flights around the congestion (there needs to be a trial flight plan mechanism for knowing what other routes are "congestion-free"). When a flight is rerouted out of a congested airspace, the airline can sub flights up as if the flight is cancelled.

- The FAA monitors the situation. If delays remain large, a GDP could be issued (using control by time of arrival?). If the delays are smaller, maybe some degree of en route metering might be a better way to solve the problem.

This approach provides each of the key elements of the GDP: the airline can assess impact and manage to its schedule, now using rerouting as another tool. The FAA has the benefit of getting airlines to voluntarily reduce the demand.

It is not clear exactly how this idea translates into the multiple constraint environments. The notion of delay equity avoids any issues of double penalties, and makes it relatively straightforward to compute the delays for a flight that is involved in multiple constraints. Independent impact assessment reports can be sent to the airlines for the multiple constraints. It would seem that because a flight can be involved in more than one constraint, the burden on the airline increases to determine the best ways to manage their schedules. However, with this burden may come additional benefits. For example, an airline might determine that by canceling a flight strategically placed in several different constraints, it could gain delay reductions for all flight involved in any one of these constraints. This bears much more thinking.

The en route weather problem is distinctly different from the sector demand problem in that during some time period, no amount of traffic is going to flow through that area. That is, reducing the demand via delays does not help. For major weather areas, it would seem that a collaborative process is necessary for determining possible routes for the major impacted traffic flows. For smaller flows or smaller weather areas, it would still be possible to generate an impact list for the airlines. It could reflect a predicted delay (the time for a short-lived weather event not longer affect the route) or simply indicate an impassable route. Airlines could generate new routes avoiding these areas. New sector congestion situations would be resolved as described above.

Another concept for the en route weather problem is the following. The worst disruptions to air traffic occur when weather has a widespread effect in an area with a high traffic volume. When it is early enough to reduce the volume in a manageable way, there is no way to predict which flights the weather is going to impact. When it is late enough to know exactly where the weather will be, there is no way to manage an orderly reduction in volume. Instead, flights are diverted or hold off-the gate, flights get trapped on the ground due to saturation of the available open routes, etc. Is it possible that using a product such as the CCFP, there could be managed reductions in volume that would provide the slack necessary to handle the weather as it evolves. Again, I think such a approach could be cast in the same terms as the GDP. Consider a scenario:

- An area over ZOB and ZID is identified as having a high probability of 25% to 50% thunderstorm coverage 6 hours from now.

- Based on historical data, the FAA notifies the users of the average delay that can be expected on the flights impacted by that area (not knowing at this time which routes are actually going to be blocked or when).

- Airlines either decide the delay is acceptable or use cancels and reroutes to reduce their demand.

- Airlines that reduce their demand are given credit (thank you Roger). This could possibly be done through a flight-by-flight substitution process.

- Later, when the weather problems are being managed, this credit is used to equalize the delay. That is, a flight that has an hour credit would get an hour less additional delay than a flight that has no credit.

Again, this concept requires much more thought.

Intent-Neutral Total Imposed Delay (INTID)

Philosophy:

Concept:

When weather or other special use conditions render a region of airspace unavailable, traffic bound for that region must be rerouted to adjacent airspace. A conflict arises between two groups of flights: A, those flights bound for the region, and B, those flights bound for (or in) the adjacent airspace. Flights from both groups must be spatially or temporally delayed to accommodate limited capacity. See Figure #.

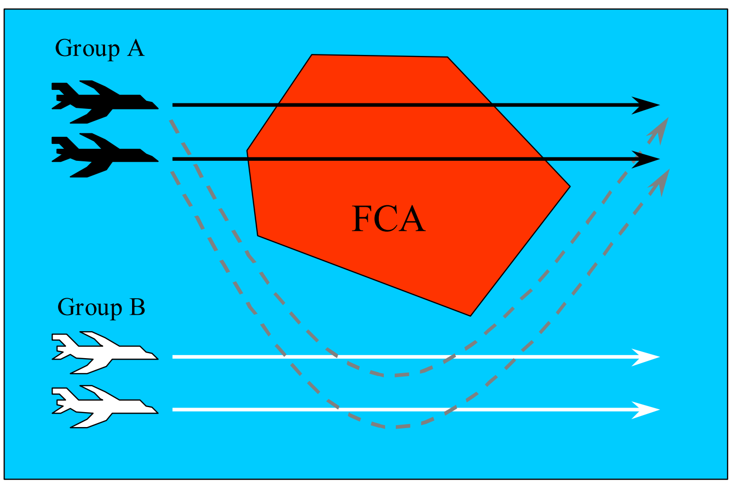

Figure #: Group A flights, bound for flow-constrained area, compete with Group B flights for adjacent airspace.

At first, it may seem reasonable to minimize the disruption to group B, on the grounds that they are innocent bystanders. This is especially true for those flights that transitioned from group A to group B by proactively rerouting around the constrained region of airspace. However, group B will be comprised of flights of varying planning horizons (short, long) and varying status (grounded, airborne). It will be difficult, if not impossible, to discern user intent based solely on filed flight plans. Therefore, as a practical matter, allocation of en route resources must be intent-neutral.

In order for a controlling authority to equitably allocate en route resources, some measure of comparative treatment must be invoked. One such measure may be imposed delay. Unlike the GDP setting, delay in the en route setting can take several forms. Of course, there is traditional ground holding, as in a GDP. This can take a more subtle form. For instance, a flight may absorb significant ground delay waiting to merge into an overhead stream of traffic. Another form of delay is spatial rerouting. Lastly, there are downstream delay effects within an air carrier's operation. Late arrival of one flight may propagate delay to subsequent flights dependent upon the crew, equipment, or passengers. This can be measured by various means, such as down line passenger delay.

These forms of delay can be partitioned into three basic categories: air delay, ground delay, and downstream delay. There are two ways to handle delay imposed on a flight by TFM initiatives. One way is to total all forms of delay into a single delay metric, total imposed delay (TID):

TID = (air delay) + (ground delay) + (down line delay)

TID could then be used as a baseline for imposing delay across flights competing for the same resources.

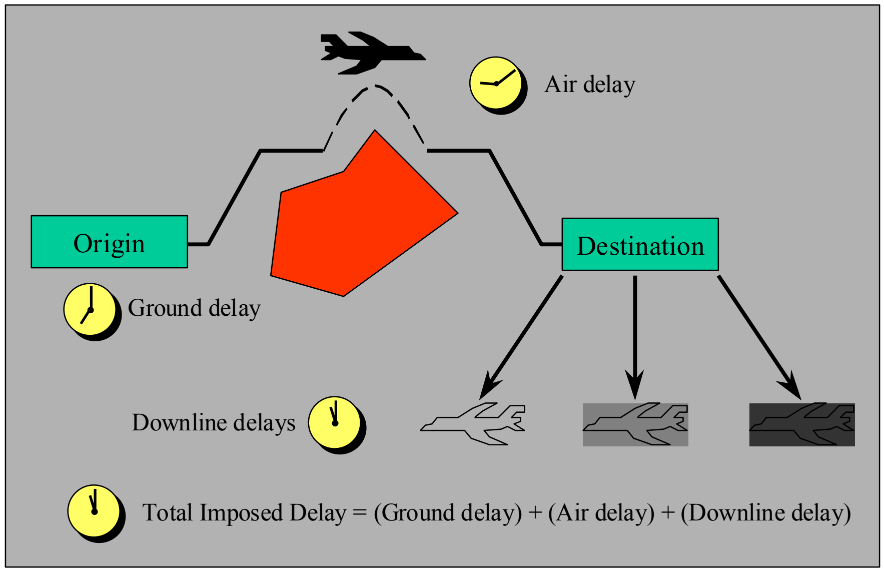

Figure #: Total imposed delay is the sum of delays at origin airport, en route, and downline delays.

Airlines are acutely aware that these forms of delay may have varying impact on their operation. The other way to measure delay is to apply a weight to each of the delay forms before summing, thus creating a weighted, total imposed delay metric (WTID):

WTID = W_g (ground delay) + W_a(air delay) + W_d (downline delay)

The problem with such a weighting scheme is that the weights would vary with the operating philosophy of the carrier. For instance, some carries value passenger comfort more than others. These weights would be highly carrier specific and even situation specific. Even the carrier may have difficulty in setting uniform weights for all of its flights. So, it seems that it would be too difficult to handle carrier preferences in initial allocation of resources. Subtleties of operation are best addressed by carrier responses to allocation of resources. Some mechanism, such as GDP substitution and cancellation procedures, would need to be developed.

This leaves us with the concept of total imposed delay as the equity metric. A scheme similar to Equity via Net Arrival Delay (ENAD) could be developed. The following features of an en route resource rationing system should be emphasized:

- Optimization of total through-put
- Equity consideration for varying categories of Users (e.g., departures merging with over-flights)
- Indifference with respect to User intentions (intent-neutral)
- Acceptance of User preferences for routes and delays
- Flight restrictions would be accessible by carrier flight planning systems

Appendix C.

Reroute Roulette

Philosophy: In a flow-constrained situation, a certain percent of the traffic must be displaced to accommodate decreased capacity. Ideally, NAS Users should be made aware of the constraints and given the opportunity to reroute around the flow constrained area. If not enough flights are moved, then traffic displacement would be performed by a random selection, or lottery, process. A list of flights intending to fly through the FCA would be formed and flights would be randomly selected from the list and rerouted until demand is in line with capacity. Those Users who voluntarily rerouted around the flow-constrained area would received some special consideration in the lottery.

Concept:

Whenever demand exceeds capacity for use of airspace, a decision must be made as to which flights will be rerouted by traffic flow management. It would be best if the carriers would draw down the demand themselves. So, it will be necessary to implement a constraint and demand notification system for enroute resources. TFM and NAS Users would share such a tool. An official notification, or advisory, would be issued warning Users of those regions of airspace that require demand reduction. Users would be given a period of time to voluntarily draw down the demand over the time horizon of concern. If not enough flights are rerouted, then TFM must implement a formal procedure for rerouting flights.

Equitable treatment of NAS Users is a prime concern for an TFM initiative. Under this proposed rationing scheme, a random selection process would maintain equitable treatment of flights. For example, suppose that a demand exceeds capacity for a fixed region of airspace. Then for each time period t of concern, X_t flights must be moved to reduce demand to the capacity level during period t. A list of flights intending to use the region would be formed, and flights will be selected one at a time from the list until X_t flights have been chosen, These flights would then be rerouted around the flow constrained area.

Slight inequities may arise in a given implementation of this “reroute roulette”, but these will be smoothed out with more frequent use of this rationing concept. In the long run, flight displacement will be in proportion to a carrier’s propensity to use flow constrained areas.

In order to provide an incentive for Users to alleviate congested situations without resortment to TFM initiatives, special consideration must be given to those Users who reroute flights in response to FCA advisories. This can be accomplished by granting a ‘free draw’ in this lottery process for each flight moved. Suppose that carrier XAL has voluntarily rerouted a flight around an FCA. Then the first time that an XAL flight is drawn from the selection pool, the flight will be returned to the pool. Similarly, if XAL voluntarily rerouted 10 flights, then the first ten times an XAL flights is drawn from the pool, the flight would be returned.

Under the current filing system, Users would retain the ability to subvert the random reroute process by refiling through the flow-constrained area. Therefore, it would be necessary to ban the selected flights from use of the resource for a sufficient period of time (e.g., 24 hours after the advisory is issued).

Issues:

- Will this work for carriers with very small numbers of flights?
- Are we sacrificing optimality by selecting flights at random?

Problem:
Only addresses spatial delay, not temporal

Appendix D.

Traffic Classes

Philosophy:

A general philosophical approach to traffic flow management and rationing involves the use of aggregate traffic classes. The use of traffic classes provides a hierarchical approach to problem solving that is consistent with many decision support tools and perhaps, more importantly, with the way traffic flow managers approach problem solving. For example, traffic flow managers might speak in terms of northeast arrivals, departures to the west, over-flights, etc. Since any automatic algorithms for controlling flights and allocating resources must be monitored and controlled by human decision makers using an approach that is consistent with their intuition has many advantages.

Concept:

We define a *traffic class* as a set of flights with certain common characteristics. The common characteristic(s) could be a route of travel, a destination point (airport, departure fix), or geographical description (e.g., eastbound flights). A flight may be in more than one traffic class. For instance, a flight could be in the eastbound traffic class and in the EWR arrival traffic class. Each flight must be assigned to at least one traffic class. Establishing an "other" (default) traffic class can insure this.

We envision that some of the traffic classes would be established automatically by route bundling or by historical traffic patterns. The user would reserve the right to delete a traffic class or to create a new one. One of the most fundamental traffic classes would be those flights that are scheduled or predicted to pass through a flow constrained area (FCA).

Traffic classes serve four purposes. They allow the user to

- aggregate a large number of flights into a manageable number of groups for the benefit of discussion and conceptualization of flows;
- insure equitable treatment for flights through explicit representation (e.g., departures attempting to merge into an overhead stream);
- control flows by entering restrictions, flow-rate objectives or capacities on each traffic class;
- reroute flights on a large scale by moving flights from one traffic class to another.

This concept is deliberately general/flexible, and is designed to aggregate flights much the same way that traffic managers do when thinking about traffic flows.

Rationing. The rationing scheme would go as follows. Once an FCA and the appropriate traffic classes have been defined, capacities must be entered for each of the traffic classes. (Default values would be effectively infinite.) If a flight is in a constrained traffic class, then there are two options for relieving the excess demand:

1. reroute (reassign to another traffic class);
2. ground delay.

This will create conflicts for resources that must be resolved. In particular, this must be done in a way that satisfies the flow constraints on all of the traffic classes. We propose that this flow balancing be done through an optimization problem (integer program), which we now describe. The input to the optimization problem would be

- a set of time periods $t = 1, 2, \ldots, T$
- a set of traffic classes, TC_1, TC_2, …

The decision variables are:

$TC_k\,(t)$ = the number of flights in the k^{th} traffic class during interval t

The integer program (IP) is multi-objective:

Obj1: Maximize throughput (the sum of all flows in all time periods)
Obj2: Minimize inequities (and perhaps other effects) among traffic classes

The constraint sets are as follows.

(1) Integer restrictions on variables TC_k (flights per time unit)
(2) Traffic class flow constraints of the form $TC_k\,(t) \leq b_{k\,t}$, where $b_{k\,t}$ is the capacity (flights per time unit) on traffic class TC_k in period t
(3) Node flow constraints of the form $TC_1\,(t) + TC_2\,(t) \leq b_{k\,t}$
(4) Conservation of flow constraints (inflow = outflow) at nodes. These must link the traffic flow classes across time periods.
(5) Inequity/equity constraints (form to be determined).

The output of the IP above is optimal flow in each traffic class, for each time period:

$R_k\,(t)$ = the optimal number of flights in the k^{th} traffic class during interval t

The output of this model produces restrictions on the flows in various classes. These aggregate level restrictions would then be converted by lower level processes into either reroutes or ground delays.

Iterative Cycle. We anticipate the possibility of significant throughput/equity tradeoffs relative to the use of this model. For example, it is quite likely that a solution that equitably balanced departures from an airport with flows in an overhead stream might require a reduction in overall throughput. The overhead associated with merging departures into an overhead stream would create this throughput reduction. Thus, it is likely that effective tools based on this model would keep a tally of both a “level of equity” and throughput. By iteratively calling the model the user would successively refine the solution until an effective balance between equity and throughput was achieved.

Appendix E.

Miscellaneous Thoughts/Concepts

Flight Sequencing.
Let us take the first and last points of the reroute as indicated by a specialist and connect them with a big circle line. Then let us draw a segment of the big circle that is perpendicular to this one through the first point of the reroute. Now, for each selected flight let us detect a point in time when the flight crosses this segment. The time when the flight is supposed to enter this point will be called *impact time*. This will be an analog to aircraft landing time (kind of an aircraft enters the FCA…). This impact time will be used for ration-by-schedule algorithms.

As we see, the flights' sequence depends on the proposed reroute. There may be other definitions, based on some kind of "averaging" the flights' original routes to get the "main direction" of the flight set that is being considered. Then it will not depend on a specific reroute, and thus may serve for multiple reroutes. This "averaging" idea definitely requires more thought, but it is clear that minor deviations in direction of the resulting line would cause only very small variations in impact time.

Bulldozer.
Now, for each flight in the list we will make a decision: to reroute the flight (to any of the so far proposed reroutes) or to delay it. This decision is based on what is faster for this flight: to wait on the ground till the FCA clears, or to fly around it. [Actually, instead of "faster" we may consider "better", where "better" does not necessarily mean "faster". For example, it may be five minutes faster to fly around, but it would require more fuel and thus does not worth it. This requires more thought, but for the first cut we may assume that "better" simply means "faster".]

This decision is made for all the flights in the list, one-by-one, according to the above-mentioned sequence. This mechanism is called "bulldozer" just to underline analogy with the GDP for arrival airport case, when new arrival times are being created sequentially according to flights' original arrival time.

As the flights become rerouted this way, they start using sectors they did not use before and may thus overload sector capacities. (The issue what sector capacity is and how it is defined is by itself rather difficult. Some discussion on it is done later in the document). If this is the case, several solutions may be offered. First, just complete rerouting the flights in the list, and then report the problem. Second, stop rerouting as soon as the overloading is observed and start using other reroute(s) (if any). Third, stop rerouting and suggest to the specialist to unload the impacted sectors, moving some flights from them further away from the FCA. Forth, fifth, and so on – remains to be defined.

Demand List.
How to characterize initial list of flights that contribute to an FCA demand? Two possibilities were considered to define the list:

1. Use early intent (EI) message if available, use historical routes otherwise (EI gives fuel efficient plan for the day and time).
2. When airspace is defined as a problem area, specialist pulls a list of flights intercepting it. This is the original demand.

Comparing the two:

- The second option is easier to implement because this data is already in ETMS. EI should be in the system by spring, but still some development will be needed for the first option.
- The first one does not reflect GA flights, while the second one does (to some extend).
- For the current data flows and DB design, if an airline cancels a flight before snap shot it does not have benefit for this. The way to fix this is to make a change to Sector DB design that would keep track of cancelled flights the same way as Airport DB does now.

Questions remain: What if an EI message was filed 12 hours ahead of time, and then was changed 2 or 4 hours later - which one to use in demand calculations? or impose a cut off time? How do you distinguish between reroute as the reaction to a problem and reroute just because an airline decided so anyway?

Rations.
To deal with en-route rationing, or ration-by-schedule in the sky, we need to consider different approach to demand definitions. First, rather then using instantaneous demand (peak demand within a 15-minute interval as done currently with M/A), we need to reflect how long flights are using the sector/FCA. Second, we need to define what portion of the resource is consumed by each flight and by a set of flights belonging to one airline.

If a flight is inside a sector or FCA for "n" minutes, then "n" minutes is the flight's contribution to the sector/FCA demand. Then total sector/FCA demand will be a sum of demands for all the flights found in sector/FCA. If several flights of the same airline are in sector/FCA, then their contribution should be combined to get the airline's demand for sector/FCA, or its share of the sector/FCA usage, the basis for rationing.

Swapping.
Can we swap based on entry time? How much slop does it provide? You do not have to keep perfect. After a number of swaps slop will be created that may be corrected through compression. However, we cannot swap a flight that spends 1 minute in a sector wit a flight that spends 20 minutes.

An idea here is for an airline to be allowed to swap two aircraft that use an FCA if as a result of this swap the usage does not increase for more then some predefined percentage (~20%?).

Another idea: after the swap took place we calculate the usage before it and after, and if it was reduced (for more then, say, 10%), an airline gets a credit. This is an "incentive" for an airline to use a sector this way.

APPENDIX F
Near-Term Weather Avoidance[2]

Philosophy:
The Near-Term equitable allocation scheme is a developing concept for rationing capacity during periods of demand/capacity imbalance due to severe weather or high volume. This concept is near-term in the sense that:

- We are developing it for rapid implementation within the current framework of the National Airspace System (NAS) and
- The concept will be able to function in the tactical Traffic Flow Management (TFM) domain (0-2 hours).

Plans are underway to expand this concept to also function in the strategic domain. Near-Term rationing has been developed to:

- Keep sector workloads under Monitor Alert Parameter levels;
- Ground delay inactive flights until the automation reroutes active flights;
- Minimize the arrival delay for each rerouted or delayed flight;
- Provide quick solutions to allow an operator to perform multiple what-if evaluations using different parameters and routes; and
- Avoid bias towards any NAS user or traffic class.

Concept:
Although weather products have yet to combine both the forecast duration and accuracy needed for this concept, the results of current weather research at MIT/Lincoln Laboratory (LL) and the National Center for Atmospheric Research (NCAR) look promising in a 2-hour time horizon. This weather forecast research, which is proceeding in parallel with the Near-Term rationing concept development, involves algorithms and data that will enable forecasts to include storm initiation, growth and decay. We anticipate that an acceptable 2-hour convective weather forecast will be available by the time that the Near-Term rationing concept can be fully developed and incorporated into an operational TFM system (i.e., Enhanced Traffic Management System [ETMS]).

Another reason for the 2-hour limitation for weather reroutes is the scarcity of early flight plan filings (e.g., 2 hours before departure) or other early intent information. Early intent information may allow for improved traffic predictions and better reroute planning. Although few airlines file flight plans 2 hours prior to departure, this may become more common in coming years.

This concept assumes that some flights will desire a reroute around severe weather, while others will thread their way through the weather. The concept does not separate aircraft from weather. Rather, it manages traffic demand in sectors adjacent to weather or other classes of Flow Constrained Areas (FCAs), in order to increase throughput and decrease delay, without exceeding available capacity. In other words, the idea is to effectively and safely utilize capacity in areas near severe weather and let the severity of the weather regulate the flow of traffic within the FCA.

Whereas other current concepts involve manually drawn FCAs, this concept utilizes a computer-generated weather forecast (e.g., NCWF) to automatically define convective weather FCAs. Therefore, the concept provides information on rapidly changing weather situations, which allows for better reroute planning.

To limit rerouting to those aircraft that will penetrate severe weather, the concept identifies four-dimensional intersections between weather FCAs and aircraft trajectories. This selective rerouting set is an improvement over today's operational tools (i.e., Mile-In-Trail [MIT], ground delays, ground stops), which have a tendency to also impact aircraft that will not penetrate the weather and therefore are not directly contributing to the congestion problem. However, users may modify the reroute set as desired. For example, users may add aircraft arriving at or departing from airports near the weather or remove aircraft requesting to fly through the weather (e.g., pathfinders).

To effectively manage the flow of traffic through sectors that are in close proximity to areas of severe convective weather, this concept seeks to simplify the rerouted traffic patterns by defining reroute corridors. Therefore, although the rerouted traffic will add increased traffic volume to these sectors, it will not necessarily add to their traffic complexity. Further, in order to maintain adequate safety levels, the concept will not allow reroutes that exceed sector-loading thresholds. By limiting both traffic complexity and volume, the concept helps to maintain acceptable levels of controller workload.

The integration of volume management and weather rerouting concepts is a first step towards an integrated set of TFM congestion management capabilities. This concept also includes a ground delay strategy for individual aircraft. Inactive flights that could be safely accommodated on reroute corridors, if their departures are delayed, may be assigned individual ground delays.

By utilizing several strategies working in concert, this concept may provide a better outcome than a concept that uses a single strategy or independently uses strategies that may act in conflict with each other. Integrated assessment of the TFM solution is also necessary to assure that multiple strategies are indeed combining well to provide favorable results.

The concept automatically assigns aircraft selected for rerouting to reroute corridors based on pre-defined decision logic (i.e., equitable allocation schemes). The current decision logic employed includes the following four steps:

1. Reroute active flights that are associated with the weather problem;
   - Use 1st come 1st serve (i.e., earliest arrival at best reroute corridor option for each flight) to determine which active flights to reroute first;
   - Of the reroute options available, select the option with the least reroute delay that does not exceed the Monitor Alert Parameter (MAP) thresholds for the sectors involved in the reroute; and
   - If all options exceed MAP thresholds, select the reroute involving the least sector overload.
2. If MAP thresholds are exceeded in Step 1, ground delay as necessary inactive flights that are contributing to the congestion problem (these are not flights associated with a weather problem).
3. If MAP thresholds are still exceeded from Step 2, apply airborne delay to aircraft with a weather problem to resolve remaining sector overload.
4. Reroute inactive flights associated with the weather problem and ground delay them if necessary;
   - Use 1st come 1st serve to determine which in active flights to reroute first;
   - Select the reroute option with the least combined reroute and ground delay that does not exceed MAP thresholds for the sectors involved in the reroute; and
   - If all available options exceed MAP thresholds, select the reroute involving the least combined reroute and ground delay.

It may not always be possible to successfully reroute all active flights. The reasons for this may include:

- Insufficient strategic planning, which would place too much burden on tactical initiatives
- Inaccurate weather forecasts
- Flights with insufficient fuel to utilize available capacity that may be some distance from the weather
- Flight arrivals/departures near the weather

Additionally, this concept allows for automatic assessment of planned reroutes. Assessment metrics include sector loading, extra air miles flown, total aircraft delay, ground delay, and throughput. If operational users find the assessment results unacceptable, they can modify and reassess the reroute plan.

This concept relies heavily on automation to quickly create, modify, and assess plans, so that time can be set aside for human collaboration and decision making.

An essential element to weather reroute planning or other congestion management is FAA and Airline Operational Center (AOC) collaboration. We will be expanding this concept to include additional decision support concepts and capabilities to assist with this collaboration process.

Lastly, the automation will electronically communicate the reroutes and ground delays to controllers and airline dispatchers for implementation. The process for electronic transmission of reroutes and delays is currently under development. Because hundreds of reroutes and delays may be involved in a reroute plan, it is important that the communication process be computer-assisted.

The process described above must take place within about twenty minutes. This is necessary because of the limited duration of accurate convective weather forecasts and the dynamic nature of this type of weather. In addition, for the same reasons, users will need to update TFM reroute plans often. As convective weather forecasts become more accurate at longer time intervals and as airlines file flight plans earlier, users may need to update reroute plans less frequently.

When users update reroute plans, they will leave intact those elements of previous reroute plans (e.g., reroutes and reroute corridors) that do not require alterations. The parts of the plan that may change include: addition of flight plans that were filed since the last TFM reroute plan, removal of existing reroute corridors that have become impacted by weather, and creation of new reroute corridors that take advantage of or react to new weather forecasts.

Evaluation Prototype:

This section illustrates the current implementation of this concept on the Collaborative Routing Coordination Tools (CRCT) concept development platform. Figure 1 shows a CRCT Traffic Display of the Chicago (ZAU), Indianapolis (ZID) and Kansas City (ZKC) ARTCCs with depictions of both Next Generation Radar (NEXRAD) data and NCWF polygons. The polygons represent detections and predictions of severe convective weather extending out at 30-minute intervals.

Once input, NCWF polygons automatically become FCAs. However, if desired, manual FCA entry is still available for identification of weather or other flow-constrained areas. In addition, we are exploring the use of other convective weather forecast products (e.g., CCFP and CIWS). Figure 2 shows the traffic predicted to penetrate convective weather FCAs. A list of these flights is also available via the Flight Selector window (Figure 3).

Figure 4 depicts the Weather Rerouting window. A traffic flow specialist, in collaboration with other users, inputs reroute corridors by using a combination of mouse clicks on the Traffic Display (Figure 5a) and parameter selections (e.g., minutes until start of plan, time duration of plan) on the Weather Rerouting window (Figure 4). Figure 5a depicts reroute corridors to the north and south of an area impacted by weather FCAs. Additionally, there are reroute corridors located in a gap between the two principal portions of the storm. In creating such a plan, users can more effectively disperse controller workload across multiple sectors. For example, a reroute corridor may stretch across several sectors, so that workload associated with merging and diverging traffic entering or departing the corridor can be more evenly distributed. Although sectors near the storm could expect a high-level of traffic, associated controller workload might be more manageable, because the additional traffic would be more structured. Also, additional reroute corridors can be defined farther north or south to further disperse controller workload and/or increase traffic throughput. We also maintain controller workload at or below safe levels by linking reroute corridor flow rates to NAS Monitor traffic thresholds (Figure 6a and 6b). In this way, we can fully utilize sector capacity without exceeding pre-determined limits.

The NAS Monitor displays in Figures 6a and 6b depict the maximum number of flights predicted to be in a given sector during a one- minute period. The example presented here uses the Indianapolis ARTCC. The vertical axis on the NAS Monitor display is divided into 24 15-minute intervals. The horizontal axis identifies both individual sectors within the ARTCC (upper number) and the maximum number of aircraft allowed to be in each sector during a one-minute period (lower number) (i.e., sector loading limit).

Once the user defines reroute corridors, the automation allocates those flights that will penetrate the weather FCAs to the reroute corridors. The Flight Selector window (Figure 3) is used to identify flights that will penetrate FCAs and allows these flights to be examined using a sortable list. Users can also employ the Flight Selector window to add to or remove flights from this list. The automation allows easy flight list modification. For example, arrivals and departures from airports near the storm or flights involving frequently used city pairs can be added to the list by simply filling in the origin/destination fields at the top of the Flight Selector window.

Next, the automation evaluates the weather reroute plan. The results of this evaluation, which takes less than a minute to perform, provide both visual and quantitative assessment criteria. An evaluation is essential to reassure NAS users that the plan will sustain high aircraft throughput around weather, reduce aircraft delays and cancellations, maintain safe levels of controller workload and ensure equity among carriers. The Traffic Display (Figure 5b) provides a visual representation of traffic predictions resulting from the plan. Additionally, Figures 6a and 6b depict the predicted change to sector loading. In Figure 6a (before the plan), sectors 30, 87-89 and 98 were predicted to exceed their sector loading thresholds during several 15-minute periods between 1330 and 1445. However, an examination of Figure 6b shows that no new NAS Monitor alerts resulted because of the plan (i.e., yellow or red boxes), demonstrating the effectiveness of linking the reroute corridor flow rate with NAS Monitor sector threshold limits. The blue shading in Figure 6b represents increased (dark blue) and decreased (light blue) sector loading. Figure 7 depicts a graphical analysis of the plan. This quantitative assessment of predicted plan performance includes arrival delays, extra miles flown, ground delays, the percent of aircraft that the plan can successfully reroute, and the percent of aircraft that cannot be included in the plan (along with the related rationale for their exclusion). By knowing how many flights cannot be included in the plan and why, a modified plan can be proposed to improve on the overall plan performance.

Issues:

- The concept's primary focus is on NAS efficiency
- It provides equity by not applying intentional bias and by striving to find solutions for more flights
- The concept has not yet developed collaborative procedures to reflect industry preferences
- Concept performance is dependent on:
  - Weather and traffic forecast accuracy
  - Other TFM strategies (e.g., previous strategic rerouting)
  - The scale of the weather problem
- Development of this concept is expanding to include:
  - Integration with strategic strategies (i.e., 2-6 hours)
  - Reduced capacity FCAs (in the current concept, FCA capacity is 100 percent reduced)
  - Other capabilities such as air delay

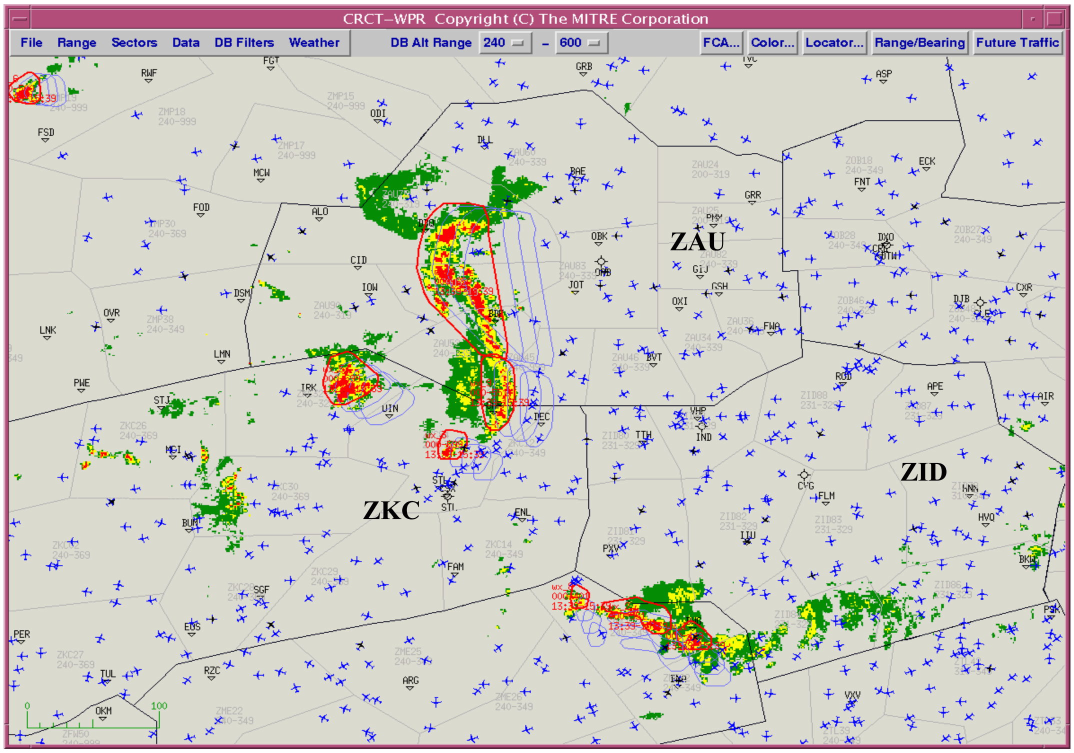

Figure 1. Traffic Display with Automated Weather FCAs

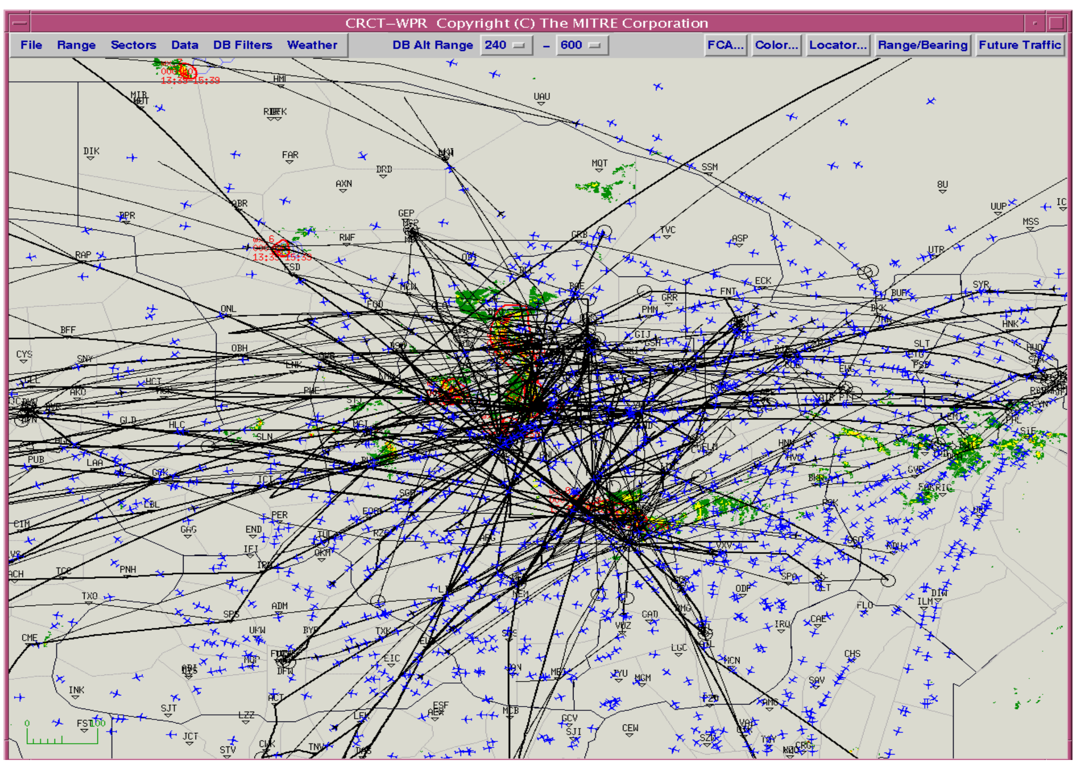

Figure 2. Flights with Trajectories Predicted to Intersect Weather FCAs (black lines)

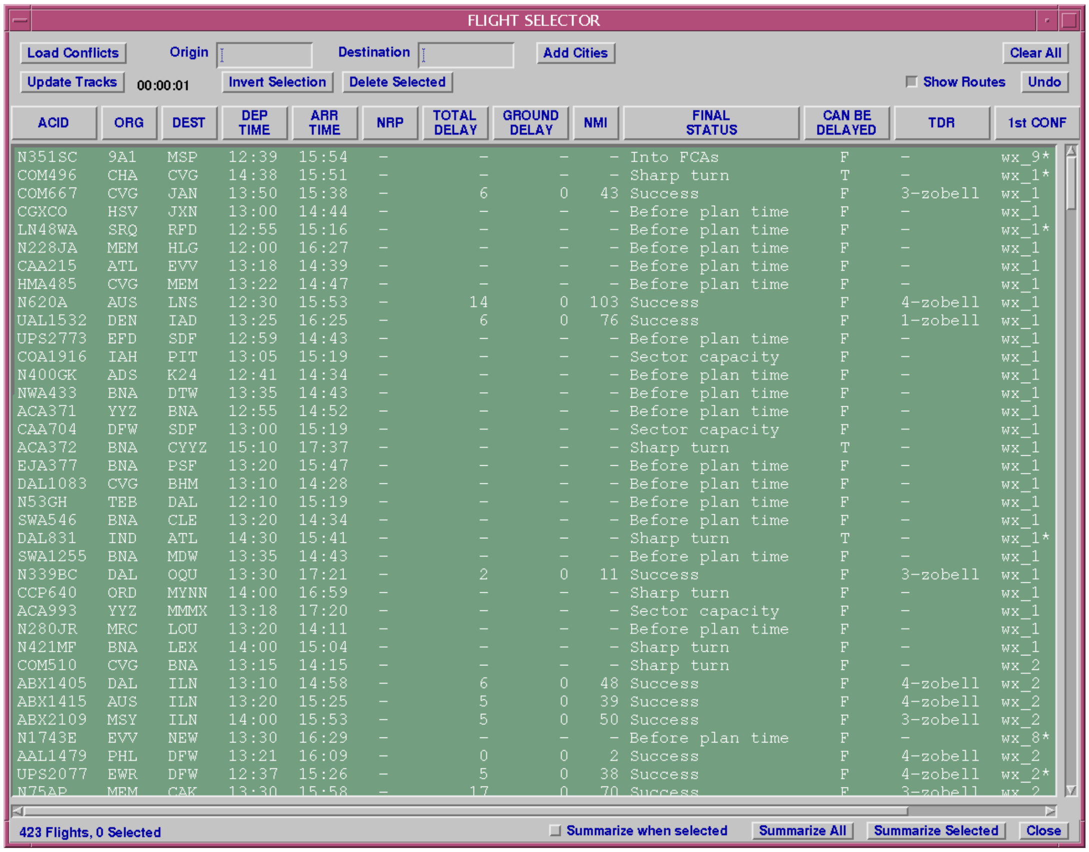

Figure 3. Flight Selector Window (Sortable Flight List)

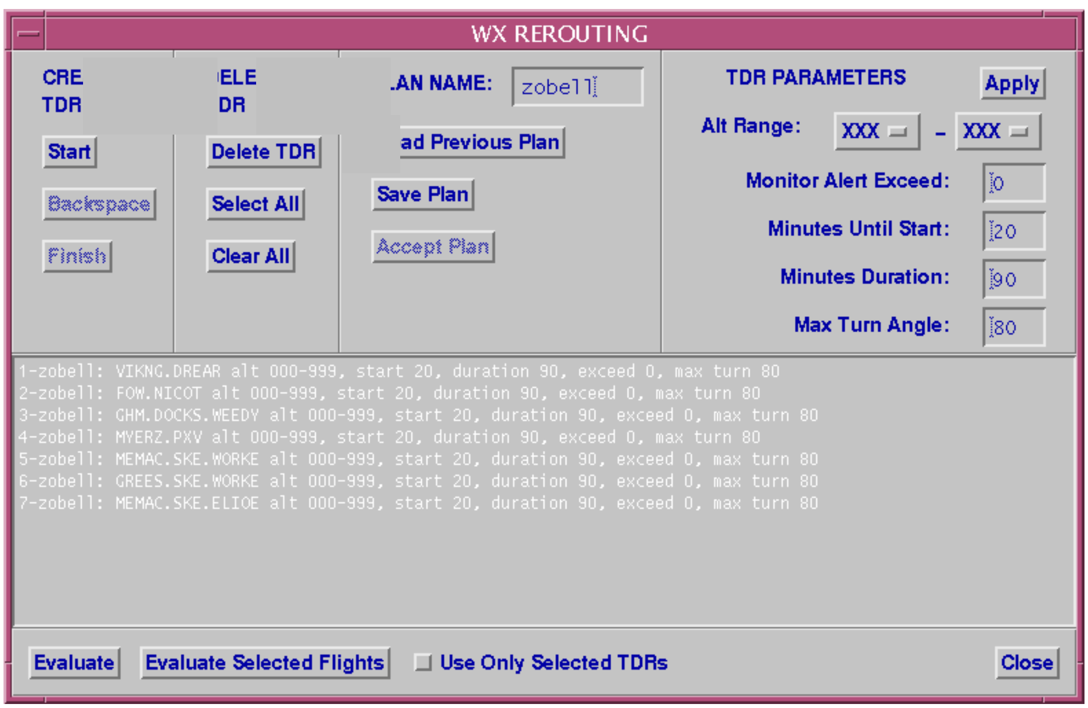

Figure 4. Weather Rerouting Window

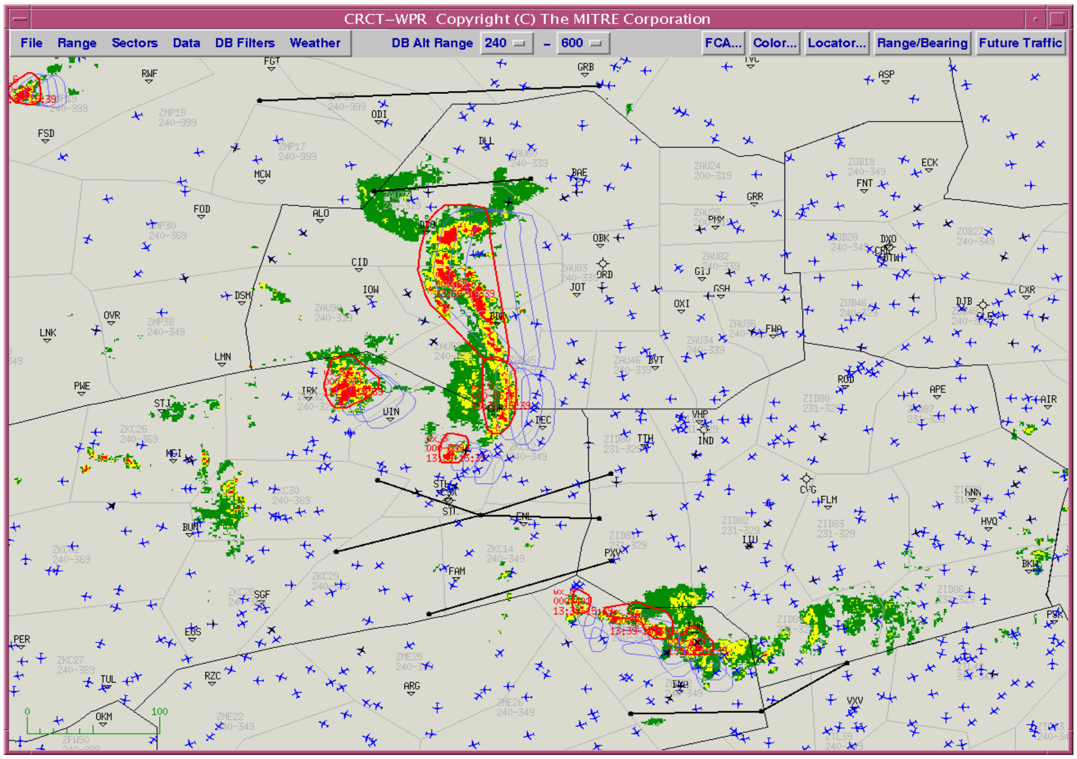

Figure 5a. Traffic Display with Weather FCAs and Reroute Corridors

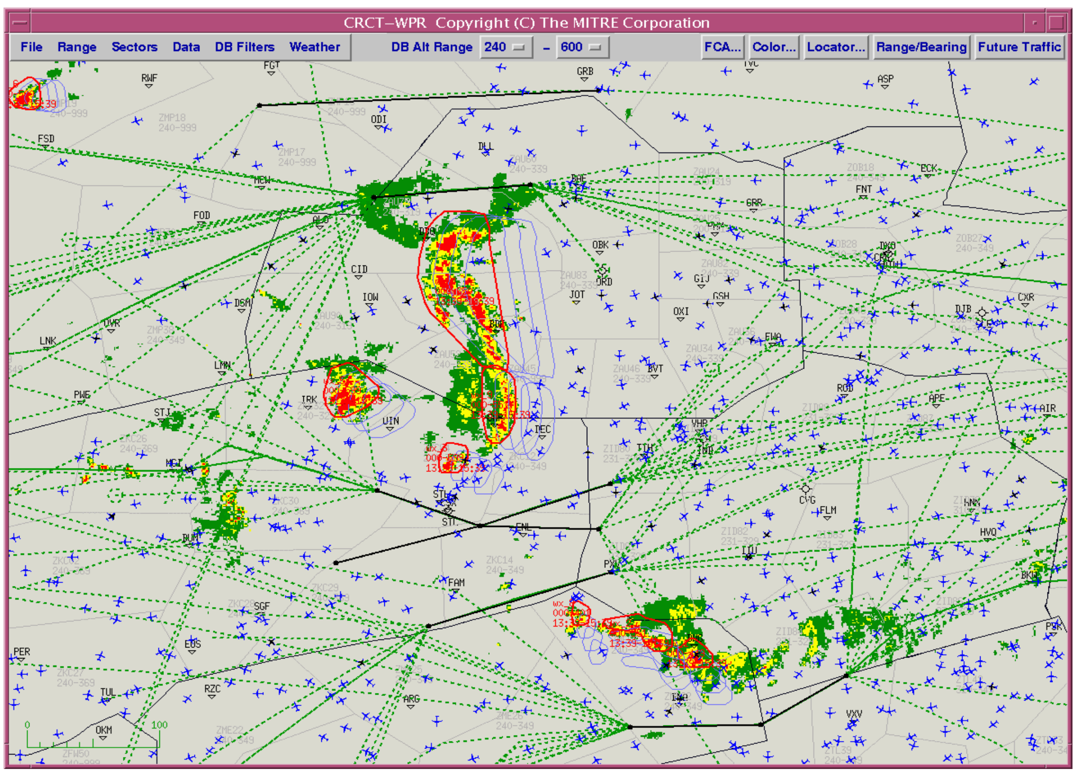

Figure 5b. Finished Plan With Reroutes Displayed (dashed green lines)

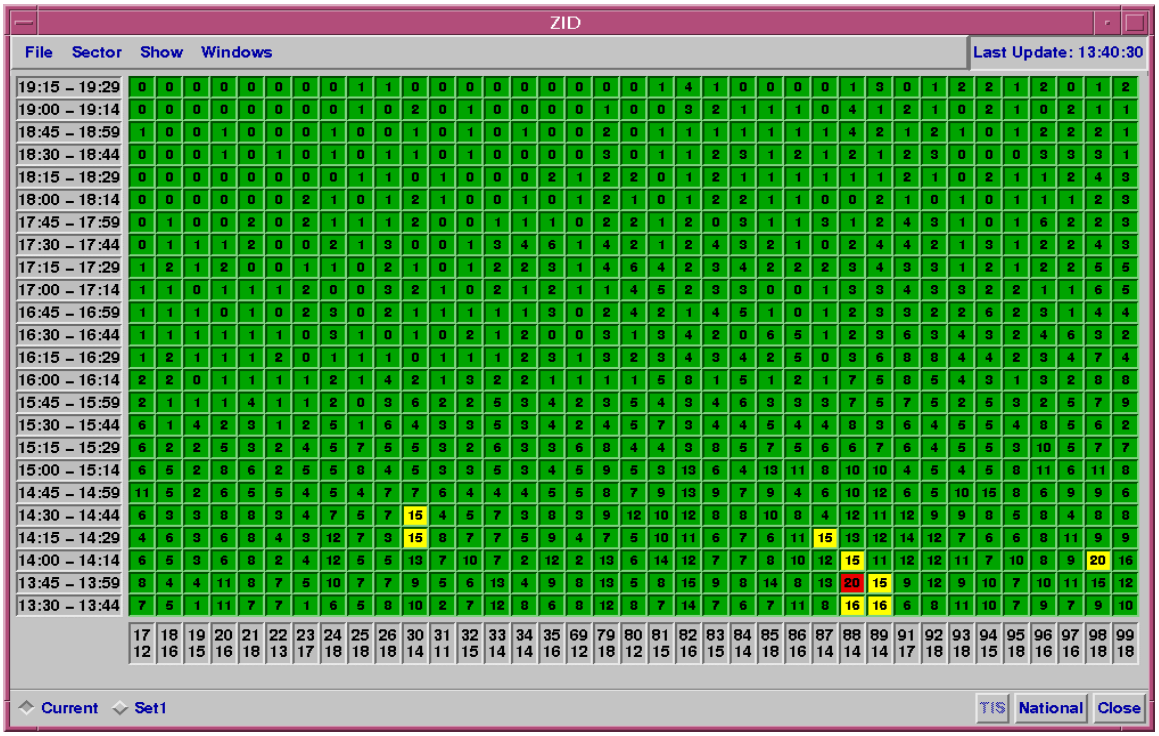

Figure 6a. Sector Count Predictions – NAS Monitor Before Plan

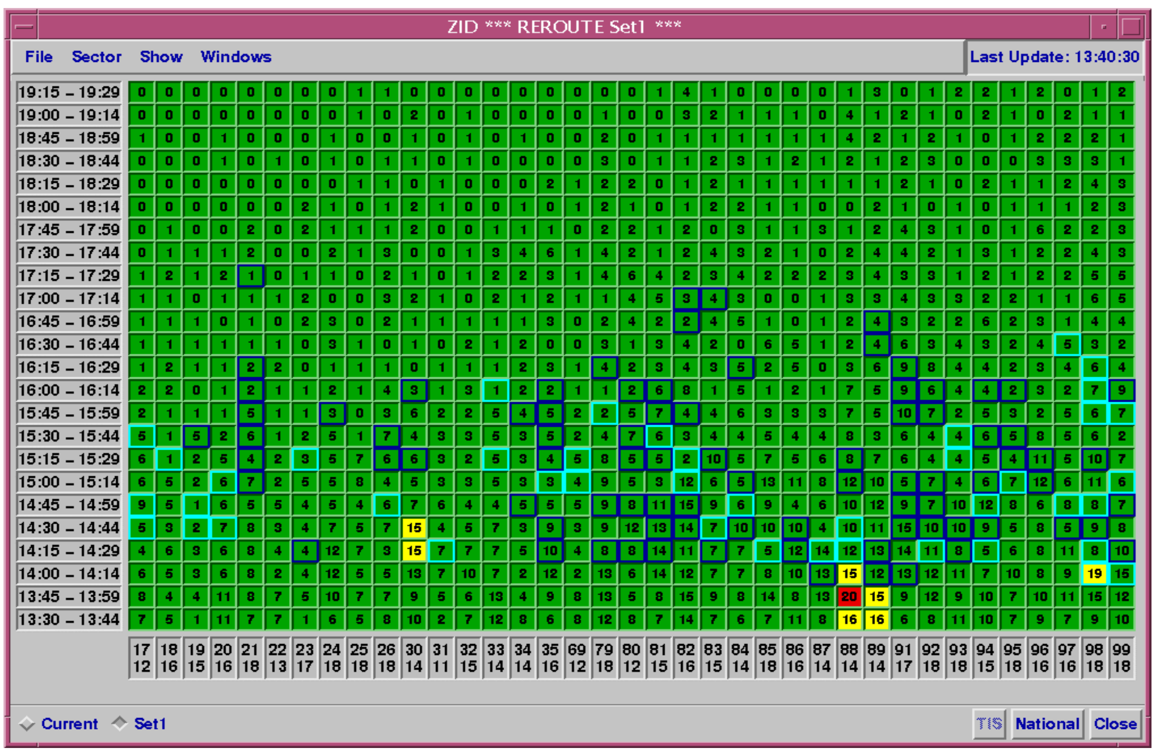

Figure 6b. Sector Count Predictions – NAS Monitor after Plan

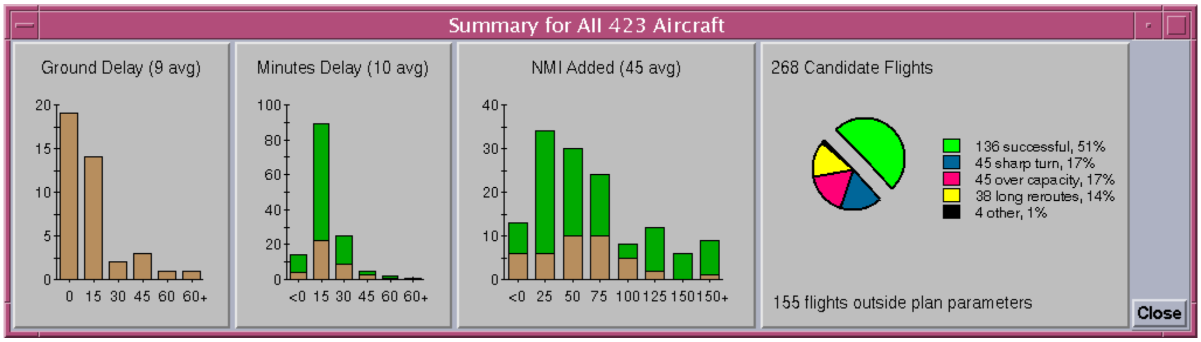

Figure 7. Plan Impact Assessment – Delays, Extra Miles Flown, and Success Summary